Trainer's notes for module 4:
Data management planning

Good practice in research data management

# Session Details

## Aims and Objectives / Learning Outcomes

By the end of this module participants will have:

* An understanding of data management fundamentals
* A practical understanding of the data management planning
* Considered reusing data and data discovery tools

## Session Topics

* Data management good practice
* Data management planning
* Reusing data

## Structure

This module is planned to be delivered via one session lasting around 100-120 minutes with a group of 12-20 attendees. (Note that you may also need to factor in a break.) The expectation is that these are postgraduate students and/or early career academics.

### Indicative timings

|  |  |
| --- | --- |
| Data management good practice | 20 minutes |
| Data Management Planning (pre activity)Activity 3: Considering the Newcastle University DMP templateData Management Planning (post activity) | 20 minutes30 minutes10 minutes |
| Reusing data  | 15 minutes |
| Session review  | 5 minutes |
|  | 100-120 minutes total |

## Set-up

Slides are provided as detailed in the "notes to accompany slidedeck". Note the use of hidden slides by default for optional activities and slides offering further detail on a topic. . (Hiding/unhiding slides is best done in 'slide sorter' view; select slides, right-click and toggle 'hide slide')

Activities are indicative rather than prescriptive. The assumption is that you are used to tailoring training outlines to meet your own needs, space and available resources! (Post-it notes, pens, flipcharts etc.)

# Notes to accompany Slidedeck for module 4

## Data Management good practice (slides 3-12)

### Slide 4: Data Management fundamentals

This section considers some of the fundamentals principles of good practice in Data Management.

This slide essentially lists a number of the key benefits of good RDM; these should be described in a positive light in advance of the (perceived?!) labours of the process of Data Management Planning in the next section.

### Slide 5: Backup, securing & archiving

Introducing the need for sound management of research data throughout the data and research lifecycle. Definitions follow.

### Slide 6: Backup and security

This slide confirms the difference between data backup and data security.

You may wish to exemplify with an example from your field or other known example.

Note the ISS guidance re. information security.

### Slide 7: Storing data

This slide provides an overview of data storage, with key considerations on the following slide.

Note the progression from short-term storage for data processing to end point storage and archiving.

You may ask attendees, "What's the default storage allowance ISS provides for staff and students?" (Answer, 4gb, see: <http://www.ncl.ac.uk/itservice/filestore/home-folder/> )

### Slide 8: Data storage: key considerations

These are the key considerations, as provided by the UK Data Archive.

Note the opening paragraphs on the linked page, and the reference to digital storage media being "inherently unreliable" and that "all file formats and physical storage media will ultimately become obsolete".

Offer an example of a media device, used for storing data, that has become obsolete in your working life (VHS tape, Zip Drive etc.) Ask attendees, what storage devices and/or file types they use for to store their data? Also, ask them to estimate when it will become obsolete?

### Slide 9: Data sharing

This is the first of three slides around data sharing; it's an underlying theme of contemporary RDM and one which some researchers are resistant to.

This slide introduces that funder requirements exist, especially for RCUK-funded research. However, the approach here is not to solely offer the big stick of RCUK-stipulation. Instead an emphasis on benefits is offered on the following slide.

As a general rule, research data should be internally discoverable and accessible within Newcastle University for non-RCUK funded research, and externally discoverable and accessible for RCUK funded research.

### Slide 10: Why share data?

You could ask attendees to offer benefits here.

Alternatively, exemplify with a selection of the following:

Benefits to researchers

* archiving provides long-term safe storage for data
* assists in implementing publishers’ data retention policies
* increases visibility of scholarly work
* may enhance researchers’ reputation
* may increase citations
* open access journal articles cited two-three times more
* enable collaborations on closely related themes, and new topics
* establish links to next generation of researchers

Benefits to the public

* production of high quality research with social value
* advance science to the benefit of society
* compliance with laws and regulations
* adoption of emerging norms – ‘open access’ publishing
* to be, and appear to be, open and accountable

Benefits to Funders

* make optimal use of publicly funded research
* avoid duplication of data collection
* maximise return for investment

Benefits to research community

* maintain professional standards of open inquiry
* maximise transparency where appropriate
* quality improvement from verification, replication and trustworthiness
* information for research design and teaching – documented methods
* promote innovation – unintended, new uses of data
* develop long time series of data

Benefits to research participants

* allow maximum use of their contributed data/information
* minimise data collection on the hard-to-reach (e.g. ill, elites)
* enable participants’ experiences to be understood as widely as ethically possible

### Slides 11: How to share data

After the "Why share data?", now on to the "how". Three broad topics are noted – repositories, licensing and clarity re. citation. Links to further detail are provided in each case

### Slides 12: Data documentation

Finally in the "good practice" section, we consider data documentation, which shouldn't be forgotten! It's one thing offering well organised data; but another vital requirement to offer the contextual information to make sense of the data and re-use it.

If time allows, an excellent way to end this section, before moving on to Data Management Planning, would be to play this short video from Louise Corti, UK Data Archive:

* How can researchers preserve, share, and re-use sensitive data? (interview, 3 mins 25 secs total, with a natural breaking point at 2 mins 20 secs)
* <http://www.sms.cam.ac.uk/media/1123202>

## Data Management Planning (slides 13-16)

### Slide 14: Data Management Plans (DMPs)

The quote on the slide is taken from the DCC website, summarising well Data Management Planning. Place emphasis as you feel appropriate.

Note to the group that, as with most planning processes, Data Management Planning can on the face of it appear to be a laborious task, but overall is an essential academic career skill and best practice, with benefits well outweighing barriers, individually and for the wider research sector, knowledge and discovery.

### Slide 15: Funder requirements

This is the first of three slides considering funder requirements.

The table, taken from the DCC website, summarises requirements from funders for DMPs. Note that it's not just RCUK funders!

Further details, including an explanation of the different columns, is on the linked DCC web page:
<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

### Slide 16: Five common themes across funders

The 5 common themes are self-evident on the slide. However, if further explanation is required around Ethics and IPR, refer to the following:

Ethics

* Balance data protection with data sharing
* Informed consent – cover current and future use
* Confidentiality – is anonymisation appropriate?
* Access control – who, what, when?

IPR

* Clarify copyright before research starts
* Consider licensing options e.g. Creative Commons

### Slide 17: What are funders looking for?

Use this slide to offer a reality check wrt what funders are looking for – they key point being that they're looking for 'reassurance' that DMP issues are in hand and the proposal can lead to "great science"!

The points on the slide are taken from the linked YouTube video which is a clip from a talk given by Peter Dukes of the MRC at the London School of Hygiene and Tropical Medicine. It's <2minutes long, and if time allows, worth playing. (Audio-only is fine – the video doesn't add anything).

### Slide 18: Why develop a DMP?

This slide balances up the funder perspective considered on the previous three slides, emphasising the benefits for the individual researcher. No need to linger on this slide if you feel the message has already sunk in!

### Slide 19: Who should write the DMP?

In module 2 we considered different roles and responsibilities generally in Research Data Management; here we consider who should be involved when writing a Data Management Plan. The emphasis is on it being a collaborative endeavour, with input from Research and Enterprise Services.

<http://www.ncl.ac.uk/res/research/index.htm>

### Slide 20: Introducing DMP Online

Now we're moving on to consider "how to create a DMP". There's a tool for that – DMP Online!

At this point, you should switch to "demo-mode" working through the DMP Online website, going through the following:

* Sign in using Newcastle University credentials
* Starting a new plan
* Adding project details
* Selecting a template
* Project overview screen
* Filling in a plan
* "My Plans"
* Exporting a plan
* Documentation

DMP Online has an intuitive and usable interface.

Further local guidance is provided on the RDM@Ncl website:
<http://research.ncl.ac.uk/rdm/tools/dmponline/>

#### DMP Online Screenshots (slides 21-30)

In case of connectivity problems, screenshots are provided in slides, hidden as default, showing the above in sequence.

### Slide 31: DCC checklist

DMP Online uses the entire DCC checklist for Data Management Planning; the whole 118 point checklist encompasses all funder requirements, grouped together to form the various templates.

Yes it's comprehensive – and long! But note that the DCC is currently reviewing this and shortening.

### Slide 32: DMP Online templates

DMP Online includes pre-defined templates for different funders, and also for individual institutions including Newcastle University! This template has been created, selecting relevant items from the BCC checklist, according to the Newcastle University draft BMP policy principles and code of good practice.

Activity 1: Considering the Newcastle University DMP template (30 minutes)

You will need enough printed copies of the handout (04 Handout DMPOnline Ncl template.docx), one for each attendee.

* Distribute the handouts and briefly describe the sections.
* Allow 5 minutes for individuals to look through the DMP template.
* After 5 minutes, encourage discussion in groups of 3-4. (The aim isn't for individuals to fill-in the DMP, rather to discuss it and related issues.) Allow 10 minutes for group discussion. Encourage groups to discuss the following questions:
	+ How comfortable are you completing the Ncl DMP template?
	+ Which sections are easy/hard to complete?
	+ Who would you need to talk to or involve in writing the plan?
* Run a 10 minute feedback session, reviewing the three questions

(Note that, if you are in a computer cluster, you could run this exercise allowing people to login to, and explore, DMP Online, referring to the guidance documentation found on the DMP@Ncl site: <http://research.ncl.ac.uk/rdm/tools/dmponline/>. However, this would undoubtedly require 15-30 minutes further time, familiarisation required with the online tool prior to discussion.)

### Slide 35: When to plan?

This slide confirms that the DMP develops throughout the research lifecycle. Make the link between the 'minimal plan' at the grant application stage and "What are funders looking for?" (slide 17).

### Slide 36: Adherence and review

This slide confirms that the DMP needs to be routinely reviewed. Why? Because funders are likely to check for adherence at the end of the project.

### Slide 37: Resourcing

Contribution to an effective DMP is time-consuming and hence costly. However, it should be costed accurately, not subsumed into other cost accounts for a project.

### Slide 38: DCC advise

The slide shows short, sharp advice from the DCC to round up this section.

## Reusing data (slides 39-44)

### Slide 40: Reusing data

A "setting the scene" slide, noting the increase in tools and services to identify and reuse existing data.

The DCC aggregation is worth clicking through to and showing if time allows.

### Slide 41: 7 Data discovery tools

Listed here are seven data discovery tools – you may wish to focus on specific ones, or add others relating to your field or that of attendees. Further detail about each tool is provided on the following slides, hidden as default; unhide as required.

1. British Library: <http://www.bl.uk/datasets>
The BL's Research Datasets programme is focused on the challenges of managing ever-increasing volumes of data, working towards its long-term preservation, and establishing standards for data citation. Datasets are [searchable](http://explore.bl.uk/primo_library/libweb/action/search.do?vid=BLVU1) through the BL's "explore" tool, and advice on finding research data is available through their [website](http://www.bl.uk/reshelp/experthelp/science/sciencetechnologymedicinecollections/researchdatasets/datasetdiscovery.html).
2. Databib: <http://databib.org/>Databib is a searchable catalog / registry / directory / bibliography of research data repositories. Individual, discipline-specific repositories are categorised by subject area. Databib is hosted by [Purdue University Libraries](http://www.lib.purdue.edu/) in the USA, and maintained by an international board of editors.
3. DataCite Metadata Search: <http://search.datacite.org/>DataCite is a not-for-profit organisation whose aim is to establish easier access to research data on the Internet. DataCite runs a service allowing data publishers to register metadata and mint [DOIs (Digital Object Identifiers)](http://www.doi.org/) for their datasets; this [metadata registry is publicly searchable](http://search.datacite.org/ui).
4. Figshare: <http://figshare.com/>Figshare is a free, personal, social web service which allows researchers to publish all of their research outputs in seconds in an easily citable, sharable and discoverable manner. All data is persistently stored online under the most liberal [Creative Commons licence](http://creativecommons.org/licenses/). All research made publicly available gets allocated a DataCite DOI at point of publication, and figshare provides a tool for browsing and searching across all data stored within it.
5. NERC data centres: <http://www.nerc.ac.uk/research/sites/data/>The Natural Environment Research Council (NERC) hosts seven separate national data centres, each specialising in a different discipline of environmental / earth sciences data. NERC issues Digital Object Identifiers (DOIs) to datasets held in its data centres. Not all of the data centres can be searched directly.
6. Registry of Research Data Repositories (re3data): <http://www.re3data.org/search/>re3data.org is another global registry of research data repositories, maintained by the [German Research Foundation](http://en.wikipedia.org/wiki/Deutsche_Forschungsgemeinschaft) (*Deutsche Forschungsgemeinschaft*) and a number of research libraries and institutions in Germany. The registry uses icons to identify important characteristics of each data repository: access and licence conditions, technical standards, *etc.*
7. UK Data Archive: <http://data-archive.ac.uk/>The UK's largest collection of digital research data in the Social Sciences and Humanities. Based at the [University of Essex](http://www.essex.ac.uk/), the UK Data Archive now includes the [Economic and Social Data Service](http://www.esds.ac.uk/) which hosts key national and international survey data and qualitative data, and the [Secure Data Service](http://securedata.data-archive.ac.uk/) which provides secure access to data that are too detailed, sensitive or confidential to be made available under standard licences.

## Session review (slides 45-47)

### Slide 46: In summary

This is a round up slide, with one summary point for each of the topics covered – you may wish to edit to align with your particular emphasis.

The Data Management Planning quote is taken from the UK Data Archive publication:
<http://data-archive.ac.uk/media/2894/managingsharing.pdf>

### Slide 47: Acknowledgements

Cited here are acknowledgements for resources used to create this module.

This module itself is licensed under a Create Commons Attribution-NonCommercial-ShareAlike license, meaning:

* You must attribute the work [[ADD HOW!!]]
* You may not use this work for commercial purposes.
* If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.